Tso (2026). *AJBSS, 1*(4)
ISSN 3083 1709 (Online)
ISSN 3083 1717 (Print)

Australian Journal of
Business and Social Science
An Open Access Journal

AI Safety Guard: Design, Prototype Implementation, and Validation Roadmap for a Privacy-Preserving Multi-Sensory Edge-AI Driver Drowsiness System

Tso, K. H.
Southern Cross University, Australia
kar.tso@scu.edu.au

Abstract

Driver drowsiness is a persistent road-safety problem whose episodic and under-reported nature complicates both prevention and measurement. This paper presents AI Safety Guard, a low-cost edge-AI prototype that combines non-contact facial-landmark analysis with bounded auditory and optional olfactory alerts. The proposed artefact uses local camera processing to estimate sustained eye closure, mouth opening and yawn patterns, and head-pose deviation; temporal decision fusion then triggers an active warning through a speaker or buzzer and, when enabled, a short, atomised scent pulse. Unlike cloud-dependent monitoring, the prototype is designed to retain no video and to record only minimal local event information. The study adopts a design-science and safety-by-design methodology: it reconstructs system requirements, specifies the hardware and inference architecture, formalises the tri-channel decision logic, and evaluates the credibility and limits of preliminary prototype evidence. Project documentation reports operation on Raspberry Pi-class hardware at approximately 10-15 frames per second, local event logging, hard-coded actuator duration, cooldown lockout, manual acknowledgement, and a scent opt-out. A website event trace reports 116 ms from a detection event to alert activation, whereas a separate pitch document claims 0.001 s actuation latency; this discrepancy is treated as an unresolved measurement issue rather than evidence of validated performance. The paper therefore distinguishes artefact feasibility from safety efficacy. It proposes a five-phase validation programme covering bench metrology, public-dataset evaluation, simulator experiments, closed-track trials, and regulatory readiness against functional-safety, safety-of-the-intended-functionality, privacy, and human-machine-interface requirements. The principal contribution is an evidence-bounded blueprint for translating a student-developed prototype into a testable driver-monitoring system while preserving privacy and explicitly managing intervention risk. The system is not positioned as a substitute for sleep, rest, or safe pull-over behaviour, but as a supplementary warning device requiring independent validation before road deployment.

Keywords: driver drowsiness detection; edge AI; facial landmarks; PERCLOS; head pose; multimodal warning; olfactory alert; privacy by design; functional safety

I. Introduction

Drowsy driving is difficult to quantify because fatigue is rarely confirmed in the same way as alcohol or speeding, and crash databases often record only indirect evidence. Nevertheless, official road-safety agencies continue to treat it as a material and under-reported hazard. The United States National Highway Traffic Safety Administration reports 644 fatalities in police-reported drowsy-driving crashes in 2024 and cautions that the precise burden is difficult to determine because drowsiness is not always observable or documented after a collision (NHTSA, 2026). In Australia, a national road-safety campaign reported that 23% of surveyed drivers had experienced a microsleep while driving (Australian Government, 2026). These figures justify research into practical detection and warning systems, but they do not imply that any single technological intervention can eliminate fatigue-related risk.

Commercial driver-monitoring systems increasingly use cameras, steering patterns, lane position, and other vehicle signals to infer attention and drowsiness. Regulation is also strengthening. Regulation (EU) 2019/2144 requires driver drowsiness and attention warning systems for relevant vehicle categories and applies to all new vehicle registrations in the European Union from July 2024.

Delegated Regulation (EU) 2021/1341 further specifies performance and privacy expectations, including closed-loop processing and data minimisation. In 2026, the United Nations Economic Commission for Europe announced a new global regulation addressing driver drowsiness and attention warning systems. These developments create both an opportunity and a constraint: low-cost retrofit systems may address older vehicles and fleets, but they must be evaluated as safety-related systems rather than consumer gadgets.

AI Safety Guard is a prototype edge device developed around a Raspberry Pi-class computer, a camera, a relay or MOSFET control circuit, an auditory alert, and an optional atomiser. The associated project documentation describes a tri-factor detection concept using eye closure, yawning, and postural telemetry, followed by simultaneous sound and scent actuation. It positions local processing and non-retention of video as central privacy features. The project received the Lead Judges' Choice Award at the 35th INCITE Awards, where it was described as a privacy-first, multi-sensory drowsiness device that detects eye closure, yawning, and head droop and triggers sound and scent (INCITE Awards, 2026). External recognition establishes novelty and social relevance, but it is not equivalent to independent scientific validation.

Figure 1. Physical AI Safety Guard prototype demonstrated in a stationary vehicle environment, showing the monitoring display, camera/compute enclosure, actuator housing, and manual control. Source: AI Safety Guard technical presentation, slide 22.

The research problem addressed in this paper is therefore not simply whether the prototype can run. It is whether its design can be specified with enough technical and evidentiary discipline to support a credible programme of validation. This requires separating three questions: detection validity, intervention effectiveness, and safety assurance. A system may accurately identify eye closure yet fail under low light; it may issue an alert quickly yet create distraction; and an olfactory stimulus may feel arousing without producing a safe, durable recovery of driving performance. Each layer requires distinct evidence.

This paper makes four contributions. First, it presents a complete reference architecture for a privacy-preserving, edge-based drowsiness detection and warning device. Second, it formalises the tri-channel fusion logic and safety controller, including persistence windows, acknowledgement, cooldown, and scent opt-out. Third, it critically appraises preliminary prototype claims and identifies unresolved measurement inconsistencies. Fourth, it proposes a staged validation and regulatory-readiness pathway covering performance, human factors, fairness, privacy, functional safety, and safety of the intended functionality.

The study is guided by three research questions: (RQ1) Can a low-cost edge-computing platform support real-time, multi-cue driver drowsiness inference while retaining video locally and avoiding cloud dependence? (RQ2) Can a bounded multi-sensory actuator architecture provide rapid warnings without creating unacceptable exposure, distraction, or repeated-actuation risks? (RQ3)

What evidence is required to progress the prototype from a demonstration artefact toward a road-ready safety system?

II. Literature Review

A. Drowsiness as an Operational Road-Safety Hazard

Driver drowsiness is a state-dependent impairment characterised by reduced vigilance, delayed reaction, attentional lapses, and episodes of microsleep. Unlike many static risk factors, fatigue can fluctuate rapidly with sleep debt, circadian timing, monotonous driving, workload, medication, and individual physiology. This variability makes one-off classification difficult and favours continuous monitoring. It also makes ground-truth labelling challenging: visible eye closure is an observable cue, but it is not identical to subjective sleepiness or neurophysiological impairment.

For system design, a warning device must distinguish between early drowsiness, transient benign behaviours, and critical loss of alertness. A natural blink should not be treated as a microsleep; normal speech should not be treated as a yawn; and a deliberate head turn should not be treated as a head drop. The operational objective is therefore temporal and contextual classification, not simple frame-level detection. Persistence windows, event frequency, multi-cue corroboration, and personal baselines are essential controls.

Official advice also limits how technological alerts should be framed. NHTSA states that coffee or energy drinks alone are not always sufficient and advises drivers who become sleepy to pull over for a short nap in a safe place. Consequently, AI Safety Guard should be described as a supplementary detection and escalation mechanism that prompts safe stopping, not as technology that makes continued fatigued driving safe.

B. Visual Indicators and Camera-Based Detection

Camera-based systems are attractive because they are non-contact and can be retrofitted without requiring the driver to wear sensors. Common features include eye aspect ratio (EAR), percentage of eyelid closure over time (PERCLOS), mouth aspect ratio or mouth opening duration, facial expression, gaze direction, and head pose. Recent peer-reviewed studies confirm that real-time visual features can support high classification performance on benchmark datasets, while also demonstrating that dataset accuracy alone does not guarantee in-vehicle robustness (Albadawi et al., 2023; Makhmudov et al., 2024).

EAR is a geometric ratio derived from eyelid landmarks. A typical formulation is shown below, where p1-p6 represent eye contour points and ||.|| denotes Euclidean distance:

$$EAR = (||p2-p6|| + ||p3-p5||) / (2\ ||p1-p4||)$$

When EAR falls below a calibrated threshold for a persistence interval, the eye channel can classify sustained closure. However, referring to a frame-level EAR rule as PERCLOS can be misleading. PERCLOS is normally a time-window statistic: the proportion of time the eyes are at least substantially closed over a defined interval. In this paper, the current implementation is therefore described as an EAR-based sustained-closure detector or a PERCLOS proxy unless a formal time-window PERCLOS computation is demonstrated.

Yawn detection commonly uses mouth landmarks, duration, and event frequency. A mouth opening during speech can resemble a yawn in a single frame, so temporal constraints are essential. The project documentation proposes classifying yawns over approximately two to eight seconds and issuing a fatigue warning when more than four events occur within fifteen minutes. These values are engineering thresholds that require empirical calibration; they should not be treated as universal physiological constants.

Head pose adds complementary information. By estimating facial landmarks in three-dimensional relation to a camera, a system can approximate pitch, yaw, and roll. Downward pitch combined with sustained eye closure may indicate a head drop more strongly than either cue alone. Edge-oriented research has shown that handcrafted feature models combining PERCLOS, mouth opening, and head pose can achieve practical model sizes and real-time operation on constrained

devices (Owen & Surantha, 2025). This supports the architectural direction of AI Safety Guard, but not its specific thresholds or accuracy.

C. Edge AI, Latency, Resilience, and Privacy

Processing camera data locally offers several advantages. First, it removes dependence on cellular coverage, which is important on regional roads and in commercial transport. Second, it reduces network-induced latency and limits exposure of sensitive video. Third, it permits a simple closed-loop control architecture in which detection directly drives a local actuator. These characteristics align with the project objective of offline operation and with the privacy direction of European DDAW regulation, which limits retention and external access to data that are unnecessary after closed-loop processing.

Edge operation also introduces constraints. Raspberry Pi-class hardware has limited compute, memory bandwidth, power, and thermal headroom. Frame rate may decline under high resolution, poor lighting, simultaneous logging, or prolonged operation. The deck reports 10-15 frames per second, which may be adequate for sustained closure and yawns if timestamps are correctly handled, but actual temporal resolution should be measured from acquisition to decision rather than inferred from nominal frame rate. Edge models must also manage camera startup, dropped frames, driver absence, occlusion, and thermal throttling.

Privacy by design requires more than a statement that video is not uploaded. A complete design should specify whether frames are stored in memory, whether debug images are written to disk, what event metadata are logged, who can access the device, and how data are deleted. Data minimisation should be demonstrable through configuration and code review. Facial landmarks can be processed without identifying a person, but the use of facial geometry does not automatically establish that the system is outside biometric or privacy regulation. Compliance depends on purpose, identifiability, storage, jurisdiction, and implementation.

D. Multi-Sensory Intervention and the Limits of Alerting

Most low-cost drowsiness prototypes use an auditory alarm. Audio is easy to implement, but drivers may habituate, mute the device, or fail to respond during severe sleepiness. AI Safety Guard adds an optional olfactory stimulus intended to provide a second sensory channel. Research on scents in driving and fatigue contexts suggests that certain odours, including peppermint, may influence perceived alertness, mood, or driving behaviour, but effects vary by scent, concentration, duration, context, and individual response (Jiang et al., 2024; Moss et al., 2023; Yang et al., 2024). This evidence supports investigation, not a claim that scent reliably reverses microsleep.

Olfactory actuation creates new risks. Drivers may have allergies, asthma, sensitivity, aversions, pregnancy-related nausea, or cultural preferences. Repeated exposure can produce habituation or discomfort. A safe design therefore requires scent opt-out, bounded pulse duration, a cooldown lockout, labelled consumables, and an audio-only fail-safe. It should also include a safety message directing the driver to pull over, rather than encouraging continued driving after stimulation. The intervention should be evaluated as a human-machine interface. A warning that is too weak may be ignored; one that is too abrupt may cause startle, steering disturbance, or distraction. Evaluation must measure not only whether the driver becomes more alert, but also lane control, reaction time, braking, perceived workload, discomfort, and false-alarm consequences.

E. Regulatory and Safety-Assurance Context

Driver monitoring sits at the intersection of product safety, automotive regulation, privacy, cybersecurity, and human factors. Regulation (EU) 2019/2144 and Commission Delegated Regulation (EU) 2021/1341 provide a useful benchmark even when a research prototype is not seeking European type approval. They require driver-drowsiness and attention warning functions for relevant vehicle categories and emphasise closed-loop processing, limited retention, warning performance, and a defined human-machine interface. The delegated regulation uses a high level of sleepiness on the

Karolinska Sleepiness Scale as a reference concept and applies technical requirements over defined vehicle speeds. In March 2026, UNECE also announced adoption of a new global regulation for driver drowsiness and attention warning, expected to enter into force by the end of 2026 (UNECE, 2026a). A retrofit device would still require jurisdiction-specific legal analysis, electrical and electromagnetic compliance, installation of safety evidence, and proof that it does not obstruct driver vision or interfere with vehicle systems.

ISO 26262 addresses functional safety for road-vehicle electrical and electronic systems. ISO 21448 addresses safety of the intended functionality (SOTIF), including unreasonable risk caused by performance limitations, foreseeable misuse, or unknown unsafe scenarios even when no component has failed. ISO/SAE 21434 extends the assurance context to cybersecurity risk across the automotive lifecycle, while ISO 24089 specifies requirements and recommendations for software-update engineering. These perspectives are complementary. A relay can fail electrically (functional safety); a vision algorithm can systematically miss a driver wearing reflective glasses (SOTIF); an attacker or unauthorised user can alter thresholds (cybersecurity); and a software update can unintentionally change timing or detection performance (update assurance). A credible development programme should therefore include hazard analysis, threat analysis, requirements traceability, verification, controlled software release, and post-deployment monitoring.

F. Ground Truth, Datasets, and Evaluation Metrics

A central challenge in drowsiness-detection research is that there is no single, universally accepted ground-truth label for "drowsy." Observable facial behaviours, subjective sleepiness, physiological changes, and driving-performance degradation do not always occur at the same time. Eye closure can indicate a normal blink, ocular irritation, or microsleep; yawning can indicate boredom, thermal discomfort, or fatigue; and head movement can reflect mirror checking or road scanning. Therefore, model evaluation based only on facial labels risks circular reasoning: the system may be declared accurate because it reproduces the same visual cues from which the labels were created.

A stronger validation design uses converging evidence. Subjective state can be measured using the Karolinska Sleepiness Scale (KSS), which was developed to capture momentary sleepiness and has been validated against electroencephalographic and behavioural variables (Åkerstedt & Gillberg, 1990; Kaida et al., 2006). Objective measures can include psychomotor vigilance, lane-position variability, steering entropy, reaction time, eye-closure duration, and, where appropriate, electroencephalography or electrooculography. In simulator studies, video-based detections should be time-aligned with these measures so that the system is evaluated against operational impairment rather than only against isolated facial expressions.

Dataset design is equally important. Public datasets provide reproducibility and enable comparison with prior work, but they may contain staged behaviours, restricted camera positions, limited demographic diversity, or laboratory lighting that differs from a vehicle cabin. Locally collected data can represent the target hardware and environment, yet small samples increase uncertainty and overfitting risk. The recommended approach is a layered dataset strategy: initial algorithm development on public data; controlled local recording for hardware-specific calibration; and prospective validation on participants and environments not used for tuning. Training, validation, and test subjects should be separated at the participant level to prevent identity leakage.

Performance reporting should go beyond overall accuracy. Drowsiness is often an imbalanced event, so accuracy can remain high even if critical episodes are missed. Sensitivity, specificity, precision, negative predictive value, false alarms per driving hour, event-level recall, time-to-detection, and confidence intervals should be reported. For a warning system, event-level timing is particularly important: the relevant question is not only whether an episode is detected but whether it is detected early enough to support a safe response. Results should be stratified by lighting, eyewear, head pose, camera distance, skin tone, and other foreseeable conditions. This discipline is necessary before a prototype can support claims about safety effectiveness.

G. Sensor Fusion, Explainability, and Personalisation

Multi-cue fusion can improve robustness because eye closure, yawning, and head pose fail in different ways. A fixed single threshold is simple, but it assumes that all drivers exhibit similar baseline facial geometry and that cabin conditions remain stable. Research systems increasingly combine multiple visual or physiological cues, while recent edge-oriented studies demonstrate that handcrafted features can remain computationally efficient on constrained devices (Chew et al., 2024; Owen & Surantha, 2025). The design question is not whether more sensors are always better; it is whether each additional cue contributes independent, interpretable information without adding unacceptable latency, privacy exposure, or failure modes.

For AI Safety Guard, explainability can be implemented at the decision level. Each warning event should identify the contributing channels, their confidence, persistence duration, and the state-machine transition that produced the alert. A driver or investigator should be able to distinguish an alert caused by sustained eye closure from one caused by repeated yawns or concurrent eye closure and head drop. This form of traceability is more useful for an early safety prototype than a single opaque probability. It supports debugging, threshold review, and incident analysis while allowing the system to retain minimal rather than raw video data.

Personalisation should be conservative and bounded. Baseline eye geometry, camera mounting, corrective lenses, and habitual head position vary across users. A short calibration sequence can estimate open-eye EAR, neutral head pose, and landmark confidence. Thresholds can then be expressed relative to baseline rather than as universal constants. However, adaptive algorithms must not silently relax safety criteria after repeated alarms. Any adaptation should be limited to documented parameters, preserve absolute safety bounds, and be resettable. Personalisation data should remain local and should not become a mechanism for biometric identification or driver profiling.

III. Methodology

A. Research Design

This study uses a design-science research approach because its primary subject is an engineered artefact. The method is appropriate when the contribution is not only explanatory knowledge but also a designed system that addresses a practical problem. The study follows five activities: problem definition; requirements derivation; artefact reconstruction and specification; preliminary demonstration analysis; and development of a rigorous evaluation pathway.

The current paper does not report a clinical trial, naturalistic driving trial, or controlled simulator experiment. It therefore avoids causal claims about crash prevention, restored alertness, or validated detection accuracy. Evidence is separated into three levels: (1) design evidence, describing what the system is intended to do; (2) prototype evidence, describing observed logs and bench demonstrations; and (3) future validation evidence, specifying what must be collected before deployment.

B. Evidence Sources

The artefact specification was reconstructed from the AI Safety Guard project website, a thirteen-page project deck, a twenty-three-slide technical presentation titled AI Drowsiness Detection System, a demonstration reference, and the public INCITE Awards announcement. The presentation provides additional photographs and diagrams covering the assembled prototype, retrofit hardware architecture, offline facial mapping, eye-closure logic, yawn and head-pose telemetry, synchronized actuation, manual override, cooldown control, local data flow, and preliminary test claims. Because these materials are project-controlled sources, they are analysed as artefact documentation rather than independent validation.

The external evidence base comprised official road-safety and regulatory sources, international standards, and recent peer-reviewed research on visual drowsiness detection, edge-computing models, dual-modal systems, and olfactory intervention. Sources were selected for relevance to the design rather than to maximise citation count.

The presentation figures reproduced in this manuscript are used to make the artefact inspectable. They document the design intent and report prototype behaviour, but they do not convert project claims into validated results. Where a slide uses absolute language such as "zero latency," "absolute privacy," or "validated," the manuscript restates the claim in evidentiary terms and specifies the audit, dataset, or metrology required to substantiate it.

C. Design Requirements

Table 1. Design requirements and required verification evidence.

ID	Requirement	Design interpretation	Verification evidence
DR1	Real-time local inference	Continuous processing on Raspberry Pi-class hardware without network dependence.	FPS distribution, dropped frames, thermal load, decision latency
DR2	Multi-cue detection	Use eye closure, yawn behaviour, and head pose with temporal persistence.	Sensitivity, specificity, event-level F1, ablation analysis
DR3	Bounded intervention	Audio alert plus optional short scent pulse, with manual acknowledgement and cooldown.	Actuator onset, pulse duration, lockout integrity, fail-safe tests
DR4	Privacy by design	No cloud upload or video retention in normal operation; minimal local event logs.	Network inspection, storage audit, data-flow review, access controls
DR5	Human control	Physical scent disable switch and acknowledgement/reset.	Usability, misuse testing, control accessibility
DR6	Safety assurance	Prevent repeated actuation and provide safe degraded modes.	FMEA/STPA, fault injection, watchdog and power-loss behaviour
DR7	Equitable performance	Test across lighting, eyewear, facial characteristics, headwear, and seating positions.	Subgroup error rates and confidence intervals

D. Analytical Approach

The analysis consists of architecture decomposition, algorithm specification, evidence reconciliation, and safety-gap analysis. Architecture decomposition maps sensing, inference, decision fusion, actuation, and logging. Algorithm specification defines observable features and temporal rules. Evidence reconciliation compares claims across project sources and flags discrepancies. Safety-gap analysis applies functional-safety and SOTIF reasoning to identify hazards, foreseeable misuse, and missing evidence.

No inferential statistics are calculated because the available material does not provide participant-level observations or a sufficiently documented test dataset. Where the project reports a false-positive rate, frame rate, or latency, the paper records the claim, its source, and the evidence limitation.

E. Design-Science Rigour and Evidence Hierarchy

Design-science research is credible when the artefact, the problem context, and the evaluation logic are explicitly connected. In this study, rigour is pursued through traceability rather than through post hoc claims. Each design requirement is linked to an identified road-safety need, a known technical limitation, or a regulatory and ethical constraint. The artefact is decomposed into sensing, inference, actuation, control, and governance layers so that claims can be assessed at the appropriate level. For example, successful GPIO activation is evidence of actuation feasibility, not evidence that the intervention improves alertness or prevents crashes.

The evidence hierarchy used in this paper distinguishes five levels. Level 1 is design coherence: the hardware and software components can plausibly implement the intended function. Level 2 is

bench feasibility: the assembled prototype executes the sensing and actuation workflow under controlled conditions. Level 3 is algorithmic validity: labelled datasets demonstrate defined detection performance across test conditions. Level 4 is human-factors efficacy: controlled participant studies show that alerts are perceivable, acceptable, and associated with improved short-term vigilance without unacceptable adverse effects. Level 5 is operational safety: closed-track and, eventually, regulated public-road evidence demonstrates safe integration, reliability, and governance. The available project material supports Levels 1 and portions of Level 2. The proposed research programme is designed to generate the higher levels without conflating them.

This hierarchy also governs language. Terms such as "prototype," "reported observation," and "proposed control" are used where evidence is preliminary. Terms such as "validated," "effective," or "road-ready" are reserved for future evidence that satisfies predefined criteria. This approach responds directly to the latency discrepancy identified in the project material: rather than selecting the faster claim, the paper requires a reproducible timing definition and instrumented measurement. Evidence-bounded reporting is not a weakness; it is a prerequisite for trustworthy translation of a student innovation into a safety research platform.

F. Visual Artefact Analysis and Figure Selection

Selected visual artefacts from the twenty-three-slide presentation were integrated into the paper to improve technical traceability. The figures show the physical prototype, bill-of-materials architecture, facial-feature extraction concept, operational threshold schematics, actuation pathway, safety lockout, privacy model, and test dashboards. Their inclusion enables reviewers to compare the written system specification with the project's implementation narrative. Nevertheless, the figures remain design evidence: numerical values displayed in the slides are reported observations or claims unless supported by independently reproducible data.

IV. AI Safety Guard System Design

A. Hardware Architecture

The documented prototype hardware stack comprises a Pi Camera V2, a Raspberry Pi 4B compute core, a 5 V relay module, a speaker, a DC atomiser, and a physical acknowledgement/reset control. Other project material describes compatibility with a Raspberry Pi 5, USB cameras, or MOSFET switching, but these are implementation alternatives rather than demonstrated components. Power conditioning, enclosure certification, mounting, and electromagnetic compatibility are not fully specified and remain future engineering requirements.

The camera is positioned to capture the driver's face without obstructing the field of view. The compute core processes frames and updates drowsiness state. The safety controller drives the audio path directly and enables the atomiser only when the scent channel is active and the cooldown has expired. Event logs should contain timestamps, channel states, and thresholds, not images.

Table 2. Prototype hardware and principal engineering risks.

Subsystem	Prototype component	Intended function	Principal engineering risks
Compute core	Raspberry Pi 4B (prototype); Pi 5 as future-compatible variant	Local OpenCV and landmark inference	Thermal throttling; storage corruption; boot failure
Vision input	Pi Camera V2 (prototype); USB camera as alternative	Driver face capture	Low light; glare; occlusion; vibration; poor mounting
Audio output	USB speaker / buzzer	Primary warning and escalation	Muted output; inaudible cabin noise; startle
Olfactory output	DC atomiser	Optional secondary sensory cue	Sensitivity; leakage; overexposure; habituation

Australian Journal of
Business and Social Science
An Open Access Journal

Subsystem	Prototype component	Intended function	Principal engineering risks
Actuation control	5 V relay module / MOSFET	Switch audio/atomiser load	Stuck-on relay; switching transient; wiring fault
Human controls	ACK/reset and scent toggle	Acknowledge warning; disable scent	Poor accessibility; accidental activation; misuse

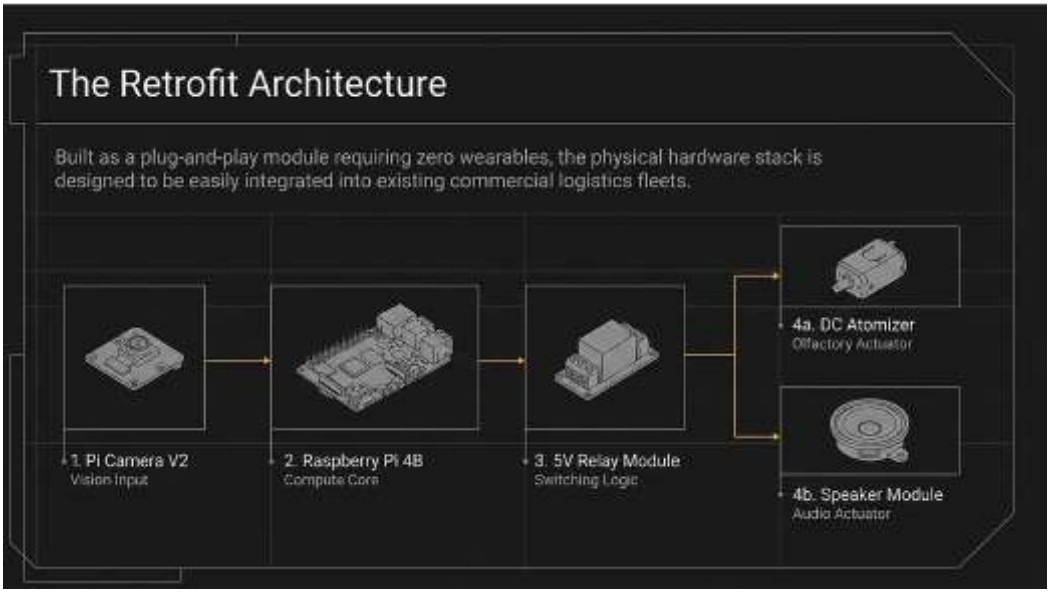

Figure 2. Documented retrofit hardware architecture: Pi Camera V2, Raspberry Pi 4B, 5 V relay module, DC atomiser, and speaker. Source: AI Safety Guard technical presentation, slide 15.

B. Vision Processing Pipeline

The vision pipeline begins with OpenCV frame capture and face localisation. Facial landmarks are then estimated using MediaPipe Face Landmarker or a dlib-compatible model. Earlier project material reports approximately 10-15 frames per second on Raspberry Pi-class hardware, whereas the later presentation states a steady 15+ frames per second. In the absence of raw frame timestamps and thermal-duration tests, this paper treats 10-15 FPS as the conservative project-reported operating range rather than a validated performance specification. For each valid face, the system derives eye landmarks, mouth contours, and selected three-dimensional facial points used for head-pose estimation.

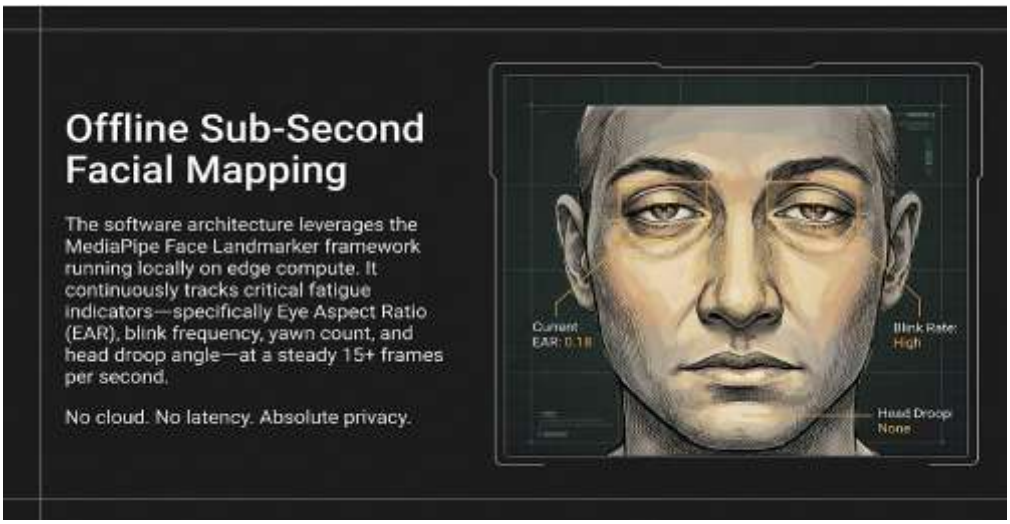

Figure 3. Presentation view of offline facial mapping and monitored indicators. The displayed 15+ FPS and "no latency" statements are project claims requiring instrumented verification. Source: AI Safety Guard technical presentation, slide 10.

The eye channel computes EAR, an eye-openness score, and a sustained-closure timer. The presentation distinguishes a normal blink from a closure episode that remains below a critical threshold for a defined persistence window. A personal baseline can be estimated during an initial calibration interval, with operational thresholds expressed relative to baseline rather than as a fixed value for every driver. The website illustrates a baseline EAR of approximately 0.32, a normal blink around 0.22, and a closure threshold around 0.15; the later presentation shows an illustrative current EAR of 0.18. These values are examples and must not be treated as validated universal thresholds.

Figure 4. Illustrative eye-openness threshold and persistence logic used to distinguish a normal blink from sustained closure. This slide is an operational schematic rather than a validated PERCLOS result. Source: AI Safety Guard technical presentation, slide 11.

The mouth channel calculates mouth opening and duration. The presentation defines a candidate yawn as mouth opening lasting approximately two to eight seconds and flags more than four yawns within a fifteen-minute window as an early-stage warning. The head channel estimates pitch, yaw, and roll through geometric mapping. A critical prototype rule combines significant head deviation with sustained eye closure exceedingly approximately two seconds. These are transparent prototype heuristics, not clinically validated cut-offs; speech, singing, coughing, mirror checks, and normal downward glances must be included in validation to quantify false detections.

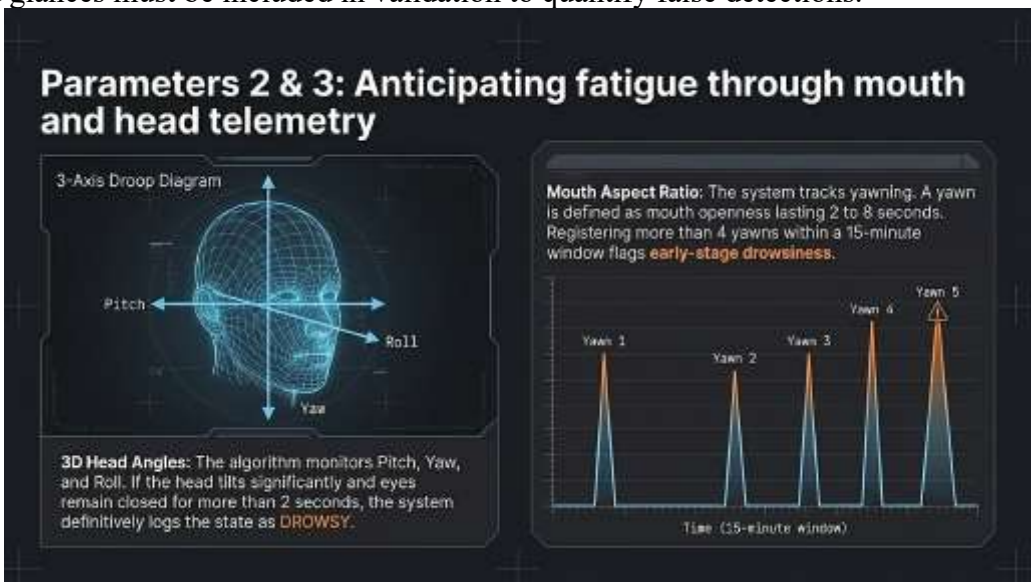

Figure 5. Prototype head-pose and yawn-frequency logic, including pitch, yaw, roll, yawn duration, and a fifteen-minute event window. Thresholds shown are project heuristics pending validation. Source: AI Safety Guard technical presentation, slide 12.

Robust operation requires explicit handling of no-face frames, multiple faces, landmark confidence, occlusion, and timestamp discontinuity. A safety-related implementation should not silently infer alertness when the camera cannot see the driver. Instead, low-confidence states should produce a system-status warning and may trigger a degraded monitoring mode.

C. Tri-Channel Temporal Decision Fusion

The prototype concept combines three channels rather than relying on blink tracking alone. Let E(t) represent sustained-closure risk, Y(t) represent yawn risk, and H(t) represent head-pose risk. A simple interpretable fusion score can be expressed as:

$$R(t) = wE\ E(t) + wY\ Y(t) + wH\ H(t) + wEH\ [E(t)\ x\ H(t)]$$

The interaction term gives additional weight to concurrent eye closure and head drop. A warning is issued only if the risk score exceeds a configured threshold for a persistence interval, or if a critical rule is met, such as prolonged eye closure combined with extreme head-pose deviation. The coefficients and thresholds must be tuned using labelled data and then frozen for evaluation.

A rules-based design is appropriate for an early prototype because it is transparent and computationally light. However, thresholds may perform differently across drivers. Future versions can incorporate adaptive baselines, calibrated probabilities, or lightweight temporal models, provided that decision logic remains explainable and testable. The system should log the reason for every warning (for example, E+H critical event) to support debugging and safety review.

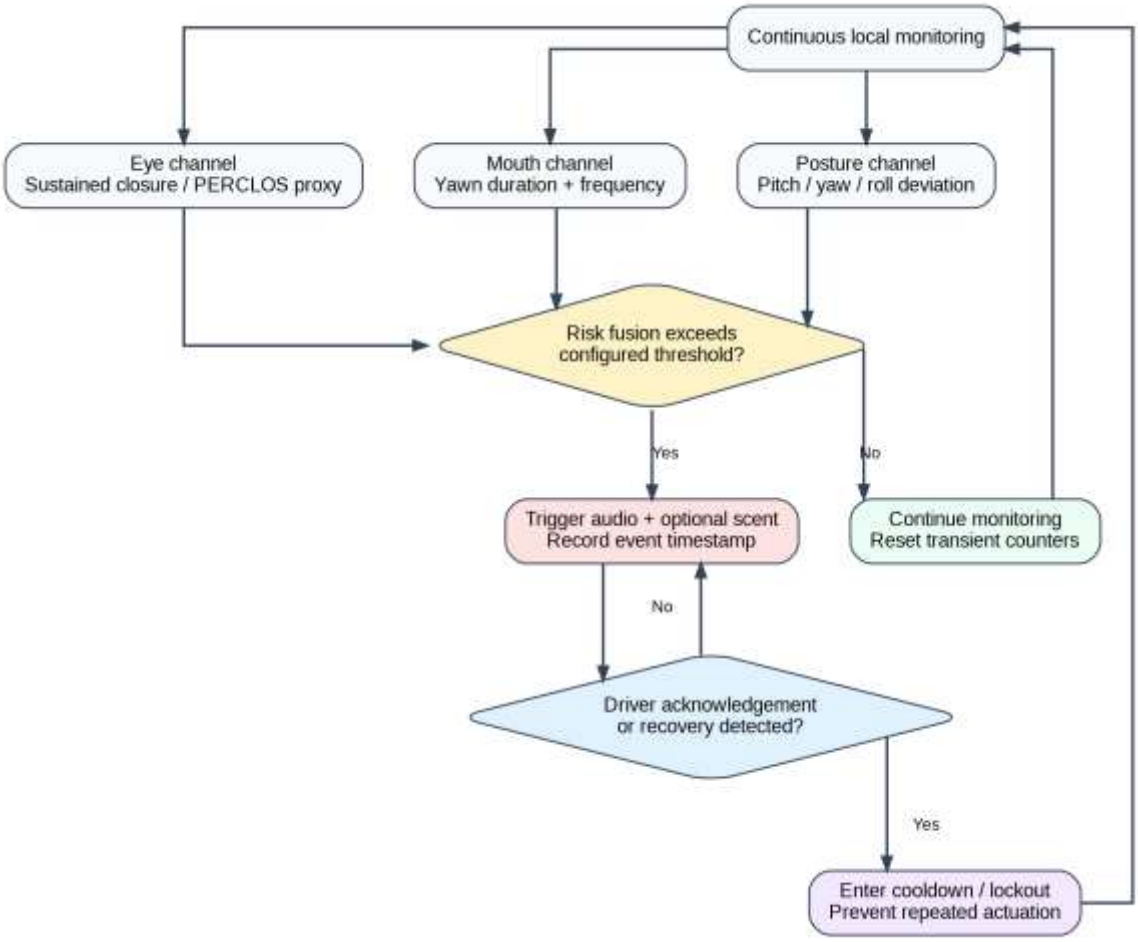

Figure 6. Proposed tri-channel decision and intervention state machine.

D. Safety Controller and Multi-Sensory Actuation

When drowsiness is detected, the audio path is activated through the local software mixer and the scent channel is commanded through a GPIO-controlled relay or MOSFET when enabled. The

presentation describes simultaneous actuation, with the alarm looping until manual acknowledgement and the atomiser limited by a short pulse and lockout. Project materials specify a maximum spray of approximately one to two seconds and a cooldown of thirty to sixty seconds. These limits reduce repeated exposure but require hardware-enforced safe states; software timing alone may be insufficient if the process crashes or the relay sticks.

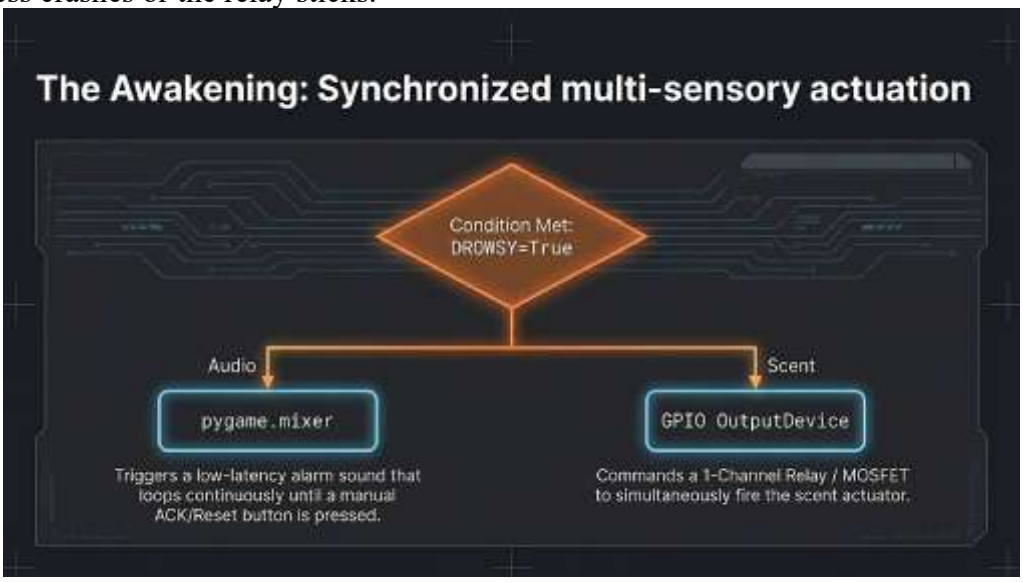

Figure 7. Synchronized decision-to-actuation pathway: local audio output and GPIO-controlled scent actuation following a DROWSY state. Source: AI Safety Guard technical presentation, slide 13.

The acknowledgement/reset button provides a human-in-the-loop control. An improved state machine should distinguish acknowledgement from recovery: pressing a button confirms that the driver noticed the alarm but does not prove alertness. The system should continue to monitor, and repeated or severe events should escalate the message to stop in a safe location. A possible escalation sequence is warning, acknowledgement request, repeated critical warning, and safe-stop instruction.

The scent option should default to disabled during initial deployment unless the driver has explicitly enabled it. Consumable composition, concentration, atomised volume, cabin ventilation, and exposure limits must be documented. The device should remain operational in audio-only mode, and failure of the atomiser must not suppress the primary warning.

Figure 8. Safe-by-design controls documented in the prototype presentation: short atomiser burst, sixty-second anti-saturation lockout, and physical acknowledgement/manual override. Source: AI Safety Guard technical presentation, slide 17.

E. Privacy and Cybersecurity Design

The privacy claim rests on local processing, volatile frame handling, and no routine video retention. The presentation depicts frames being processed on the Raspberry Pi, geometric landmarks being extracted, and images being discarded without an external network path. A production implementation must verify this claim through architecture and audit: use volatile frame buffers; disable debug recording by default; encrypt minimal logs; rotate logs; expose no remote camera stream; and confirm network behaviour through packet capture. Network interfaces not required for operation should be disabled or firewalled. Software updates should be signed, and configuration changes should be auditable.

Figure 9. Project-declared local-processing data flow. The claim of complete privacy requires confirmation through code review, storage inspection, and packet-capture testing. Source: AI Safety Guard technical presentation, slide 18.

Cybersecurity is safety-relevant because an attacker or accidental configuration change could disable warnings, alter thresholds, activate the atomiser, or access camera data. Threat modelling should cover physical access, default credentials, supply-chain dependencies, malicious updates, and denial of service. A watchdog should restart the monitoring service and signal degraded status when inference fails.

F. Calibration, Confidence Management, and Personalisation

A production-oriented implementation requires an explicit calibration workflow. At system start, the camera should verify that a single face is present within the expected region, illumination is sufficient, and landmark confidence is stable. The driver can then be asked to maintain a neutral forward gaze and perform several natural blinks. From this sequence, the device can estimate a robust open-eye EAR baseline, a neutral head-pose vector, and expected landmark variance. Calibration should fail safely when the face is obscured, multiple faces are detected, or the driver cannot complete the sequence without diverting attention. The system should allow recalibration after camera movement or a change in driver, but it should not demand frequent interaction while the vehicle is moving.

Confidence management is necessary because facial landmarks are not equally reliable in every frame. Instead of treating every output as valid, the pipeline should propagate landmark confidence into the temporal logic. Low-confidence observations should not be interpreted as closed eyes or abnormal head pose; they should transition the system into a degraded-visibility state. Short gaps can be bridged using bounded temporal continuity, whereas persistent low confidence should generate a distinct “monitoring unavailable” notification rather than a drowsiness alarm. This distinction reduces false positives and prevents the dangerous assumption that silence means the driver is alert.

Personalisation should use relative and interpretable parameters. An EAR threshold can be calculated as a proportion of the calibrated open-eye baseline, while head-pose limits can be defined around the neutral vector. Yawn logic should remain based on duration and frequency rather than mouth size alone. All personalised parameters should be stored locally, versioned, and accompanied by absolute lower or upper bounds established during validation. A user interface should expose calibration status and allow a reset to safe defaults. It should not expose complex threshold tuning to ordinary drivers, because inappropriate configuration could defeat the intended safety function.

The design should also consider driver changeovers in fleet environments. A device can offer a quick recalibration mode without creating a persistent biometric profile. The safest privacy-oriented option is ephemeral calibration: parameters are associated with the current session and discarded at shutdown unless the organisation has a lawful, transparent reason to retain them. If persistence is enabled, storage should use non-identifying device profiles and access controls. This balances usability with the project's principle that video and facial imagery remain local and transient.

G. Degraded Modes, Fault Handling, and Human-Machine Interface Escalation

Safety-related prototypes require defined behaviour when components are unavailable. A simple application that stops processing after a camera exception may fail silently. AI Safety Guard should therefore supervise the camera stream, landmark model, timing loop, audio output, relay state, and acknowledgement input. Watchdog timers can detect stalled processing, and startup self-tests can verify speaker and relay continuity without releasing scent. Faults should be classified by whether monitoring can continue, whether only a subset of cues remains available, and whether the device must disable actuation.

A proposed degraded-mode hierarchy is as follows. In Normal mode, all visual channels and outputs are available. In Vision-Degraded mode, landmark confidence or illumination is insufficient; drowsiness inference is suspended and the driver receives a non-urgent monitoring-unavailable notification. In Audio-Only mode, scent is disabled by user choice, consumable absence, relay fault, or cooldown, while the audio warning remains available. In Actuation-Fault mode, detection may continue for logging and diagnostic purposes, but the interface clearly indicates that the warning output is unavailable. In Safe Shutdown, the controller de-energises the relay and stops inference because of thermal, power, integrity, or unrecoverable software faults.

The human-machine interface should communicate urgency without creating confusion. Routine status, monitoring degradation, drowsiness warning, and critical events should have distinct acoustic or visual patterns. A critical alert should not be a prolonged startling noise that encourages unsafe manual interaction. The acknowledgement button should confirm that the alert was perceived, but it should not be interpreted as evidence of restored alertness. If high-risk cues persist after acknowledgement, the warning should recur after a bounded interval, and the interface should advise the driver to stop in a safe location.

False alarms and missed alarms require different controls. False alarms can cause annoyance, distrust, deactivation, or sudden unsafe reactions. Missed alarms can leave impairment unaddressed. A conservative design does not simply maximise sensitivity; it defines acceptable event-level trade-offs for the intended context. Fleet trials should monitor alarm frequency, driver response, deactivation, and near-miss reports. Threshold changes should be reviewed through configuration management rather than applied informally. This operating discipline makes the state machine an essential safety component rather than coding convenience.

4.8 Power, Thermal, Mechanical, and Installation Engineering

Edge computation inside a vehicle introduces environmental demands that are not represented by a desktop demonstration. Raspberry Pi-class hardware can throttle under sustained load, particularly in a sun-heated cabin. Thermal throttling can reduce frame rate and increase detection latency without an obvious software failure. The enclosure therefore requires validated ventilation or heat sinking,

temperature monitoring, and a thermal-degraded mode. Bench testing should cover cold start, hot soak, prolonged operation, supply-voltage variation, and restart after transient power loss.

Power design should protect both the device and vehicle. Automotive electrical systems experience cranking voltage drops, load dumps, electrical noise, and accidental reverse polarity. A research prototype powered by a standard USB adapter is not evidence of automotive electrical compatibility. A deployment design should include fused input, regulated supply, over-voltage and reverse-polarity protection, and separation between computation and actuator power. The atomiser circuit should default de-energised, with hardware limits preventing continuous activation even if software fails.

Mechanical installation affects perception accuracy and road safety. Camera vibration, dashboard flex, lens contamination, and gradual mount movement can alter calibration. The unit must not obstruct the driver's forward view, interfere with airbags, create sharp impact hazards, or become a projectile. Cable routes must avoid pedals and steering controls. Installation instructions should define acceptable camera zones, viewing angles, fasteners, and inspection intervals. These requirements are especially important for a retrofit product intended for heterogeneous fleets, where vehicle geometry and operating conditions vary widely.

V. Preliminary Prototype Evidence and Evidence Boundaries

A. Reported Observations

The available project material supports a feasibility claim: the assembled prototype can capture facial data locally, compute drowsiness-related features, trigger audio and relay outputs, and apply basic safety controls. The newer presentation includes an in-vehicle prototype photograph, an event-log timing example, a sixty-second lockout illustration, and a dashboard summarising six or more environmental runs. Figures 10 and 11 reproduce the latency and reliability displays as project-reported evidence. These artefacts do not establish validated on-road efficacy. Table 3 separates each reported observation from its evidentiary limitation.

Table 3. Preliminary prototype observations and their evidence boundaries

Measure	Reported observation	Source	Evidence boundary
Frame processing	10-15 FPS in earlier material; 15+ FPS in later presentation	Project deck, website, and presentation	No raw timing distribution, resolution, thermal duration, or dropped-frame statistics
Environmental testing	At least six reported runs across daylight, low light, glare, shadows, night, and cabin variance	Project presentation	Test protocol, sample size per condition, and pass/fail criteria not supplied
False positives	Claimed <1%	Earlier project deck	Denominator and event definition not supplied; cannot be independently reproduced
Event-to-alert latency	116 ms in a website example log	Project website and later presentation	Single trace; clock source and endpoint definition not documented
Actuation latency	0.001 s	Earlier project deck	Conflicts with 116 ms trace and may refer only to relay switching, not end-to-end response
Safety controls	1-2 s maximum pulse, 30-60 s cooldown/lockout, scent opt-out, and ACK/reset	Project deck and presentation	Hardware enforcement and fault-injection results not supplied
Privacy	No cloud upload; video discarded locally	Website and project presentations	Requires network/storage audit and code inspection

Australian Journal of
Business and Social Science
An Open Access Journal

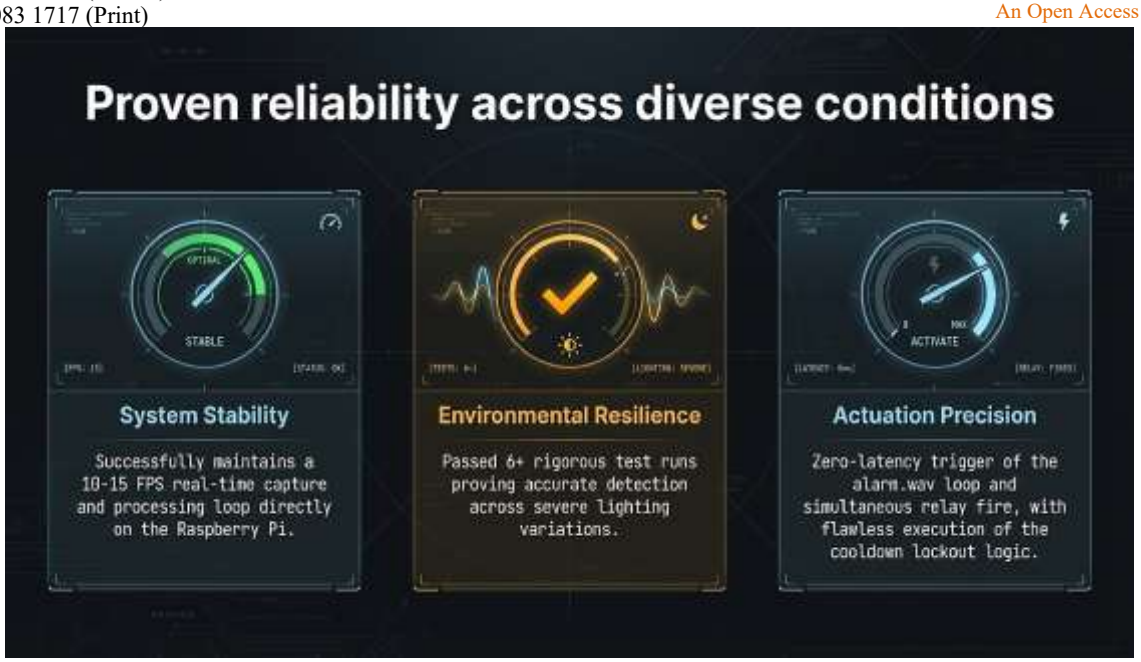

Figure 10. Project presentation dashboard summarising reported system stability, environmental resilience, and actuation precision. These are preliminary claims without supplied raw datasets or pass/fail protocols. Source: AI Safety Guard technical presentation, slide 19.

B. Reconciliation of the Latency Discrepancy

The most important internal inconsistency concerns latency. The project website and the newer twenty-three-slide presentation both display an example event sequence reporting approximately 116 ms from a detection event to actuator firing. An earlier project deck reports an actuation latency of 0.001 s. Both values can be technically plausible if they refer to different intervals: one millisecond may describe electrical switching after a GPIO command, whereas 116 ms may describe a broader software-to-hardware interval. The documentation does not define the endpoints sufficiently to determine this. The newer presentation therefore improves internal consistency with the website but does not yet provide a reproducible latency distribution.

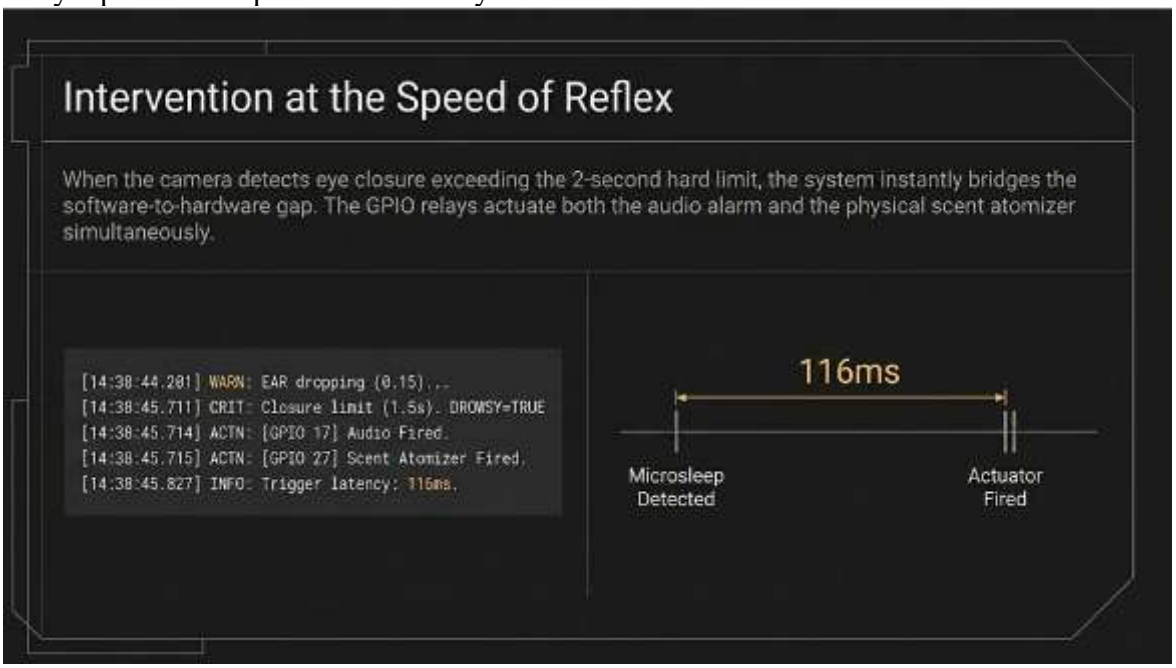

Figure 11. Example event log reporting 116 ms from microsleep detection to actuator firing. The clock source, endpoint definition, sample count, and latency distribution are not documented. Source: AI Safety Guard technical presentation, slide 16.

The correct response is not to choose the more favourable value. A repeatable measurement protocol should place timestamps at camera acquisition, landmark completion, decision threshold crossing, GPIO command, electrical output, and perceptible actuator onset. At least several hundred events should be measured under different CPU loads and lighting conditions. Results should report median, 95th percentile, maximum, and timing uncertainty. Until this is completed, the paper characterises the system as low-latency by design but not latency-validated.

C. What the Current Evidence Does and Does Not Establish

The current evidence establishes that a coherent physical prototype has been assembled and demonstrated in a stationary vehicle context. The hardware photograph, wiring architecture, interface display, and event log are consistent with the described edge-computing workflow. The hardware and software choices are also consistent with recent edge-computing research, and the privacy architecture is directionally aligned with closed-loop regulatory principles. Award recognition supports innovation and public relevance, but neither the presentation nor the award constitutes independent scientific validation.

The evidence does not establish sensitivity, specificity, negative predictive value, subgroup fairness, long-term reliability, intervention efficacy, crash reduction, or regulatory compliance. It also does not establish that scent restores alertness or that drivers can safely continue after an alert. These limitations are not defects in a student prototype; they define the next research programme.

VI. Safety, Ethical, and Regulatory Analysis

A. Preliminary Hazard Analysis

Table 4. Preliminary hazard analysis for the prototype.

ID	Hazard	Potential consequence	Proposed controls
H1	False negative during genuine drowsiness	No warning; crash risk remains	Multi-cue fusion, confidence monitoring, degraded-state warning, validation across conditions
H2	False positive alarm	Startle, distraction, loss of trust, device disablement	Persistence windows, threshold calibration, graduated audio, human-factors testing
H3	Atomiser stuck on or repeated pulses	Overexposure, discomfort, visibility or respiratory concerns	Hardware maximum-on timer, normally-off relay, independent cooldown, physical scent switch
H4	Camera obstruction or poor lighting	System assumes alertness without valid observation	Landmark confidence monitoring, IR-capable camera option, visible degraded-status warning
H5	Algorithmic subgroup bias	Unequal protection across drivers	Diverse datasets, subgroup metrics, eyewear/headwear tests, threshold personalisation
H6	Driver relies on alert to continue driving	Risk compensation and delayed rest	Explicit safe-stop messaging, repeated-event escalation, user education
H7	Privacy compromise	Exposure of face video or identifiable telemetry	No recording by default, disabled remote stream, encryption, access control, audit
H8	Power or software failure	Silent loss of monitoring	Watchdog, startup self-test, health indicator, safe degraded mode

B. Functional Safety and SOTIF

Functional-safety analysis should define the item boundary, operating modes, hazards, safety goals, and technical safety requirements. For a retrofit prototype, a practical first step is a failure mode

and effects analysis covering camera, compute service, GPIO, relay, atomiser, speaker, and controls. Fault injection should verify stuck-on, stuck-off, camera disconnect, low voltage, storage failure, and process crash.

SOTIF analysis addresses situations in which components operate as designed but the function is insufficient. Examples include dark sunglasses, facial occlusion, extreme sunlight, non-standard seating positions, deliberate head movement, speech mistaken for yawning, and rare facial geometry. Scenario-based testing should identify known unsafe scenarios, known safe scenarios, and unknown scenarios for exploration. Release criteria should focus on risk, not headline accuracy.

C. Human Factors and Ethical Use

A safety warning is an interaction, not merely an output. The alert must be perceivable in cabin noise, understandable without visual attention, and neither excessively startling nor easy to ignore. The scent channel requires informed opt-in and clear labelling. An ethics review is required before exposing participants to induced fatigue, simulator driving, or atomised substances.

The system should avoid overclaiming. Marketing language such as instant wake-up or elimination of driver fatigue is not supported by current evidence. Ethical communication should state that drowsiness can recur, alerts are supplementary, and the appropriate response is to stop and rest. Fleet deployment should not be used to discipline workers solely on algorithmic events without verification and due process.

D. Regulatory Readiness

The EU DDAW framework provides relevant design signals: operation over a defined speed range, warning performance tied to demonstrable drowsiness, closed-loop processing, and limitations on data retention. AI Safety Guard aligns directionally with local processing and immediate warning. However, a retrofit prototype is not automatically compliant, type-approved, or safe for installation in production vehicles.

Regulatory readiness would require documented intended use, installation specification, risk classification, electrical safety, electromagnetic compatibility, materials assessment for scent, cybersecurity controls, human-machine-interface validation, software lifecycle evidence, and post-market monitoring. Australian deployment would also require analysis of vehicle standards, workplace health and safety obligations for fleets, privacy law, and product liability.

E. Privacy, Data Governance, and Proportionality

The project's local-processing model is a strong privacy starting point because it reduces the need to transmit identifiable facial imagery. However, privacy by design requires verification of data flows, not only a statement of intent. A data inventory should identify every item created by the device: video frames in memory, facial landmarks, calibration values, event timestamps, device identifiers, diagnostic logs, software-version data, and any fleet-facing records. For each item, the design should specify purpose, retention, access, and deletion. Raw frames should be overwritten immediately after feature extraction unless a separately approved research protocol requires temporary recording.

Data minimisation should also apply to logs. A safety event can usually be documented with time, channel states, confidence values, threshold version, actuation state, and acknowledgement status. It does not require a face image or continuous behavioural profile. If fleet dashboards are introduced, the system should avoid turning fatigue alerts into a general worker-surveillance stream. Employers should have a clear safety purpose, consult affected workers, limit secondary use, and define who can view individual events. Aggregated safety analytics may often be sufficient for programme improvement.

Proportionality is especially important because driver monitoring occurs in a confined workplace or private vehicle. The driver should know what is sensed, what is retained, and how alerts are used. A physical scent opt-out is a meaningful control, but the same principle should extend to data. The interface should provide accessible privacy information and should not imply that video is

"never recorded" unless the software, operating system, crash diagnostics, and support processes have all been examined for unintended capture. Security updates and support logs should preserve the same minimisation commitments.

Research studies create additional obligations. Simulator or closed-track participants must receive informed consent describing fatigue induction, camera monitoring, scent exposure, withdrawal rights, and adverse-event procedures. Ethics approval should address participants with respiratory conditions, scent sensitivities, pregnancy-related nausea, or other vulnerabilities. Data collected for validation should be separated from future commercial or fleet use, de-identified where possible, and retained only for a defined research period.

F. Cybersecurity and Software-Lifecycle Assurance

Cybersecurity can become a direct safety issue when software controls alarms and a physical atomiser. Threats include unauthorised remote access, malicious threshold changes, replacement of a model or prompt file, suppression of alerts, repeated actuation, leakage of driver data, and denial of service. Although the current prototype is described as offline, offline operation does not remove threats arising from maintenance ports, removable media, local wireless services, default credentials, supply-chain components, or physical access. The attack surface should be documented through threat analysis and risk assessment aligned, where applicable, with ISO/SAE 21434.

Basic hardening should include a minimal operating-system image, disabled unused services, unique credentials, least-privilege execution, signed application packages, protected configuration files, and integrity checking at startup. The sensing application should not run with unrestricted administrator privileges. Physical interfaces should be secured or disabled after deployment, and diagnostic access should require authentication. Logs should record security-relevant configuration changes without storing unnecessary personal data. A secure recovery method is needed so that a failed update does not leave the device in an ambiguous partially operating state.

Software updates require governance because performance can change even when source code changes appear minor. ISO 24089 provides a relevant lifecycle framework for software-update engineering. Each release should have a version identifier, change description, test evidence, compatibility statement, rollback procedure, and approval record. Updates that affect landmark models, thresholds, timing, relay behaviour, or user warnings should trigger regression testing. A field device should verify update authenticity and integrity before installation, retain a known-good version, and fail to a safe state if validation fails.

UNECE Regulations No. 155 and No. 156 establish broader automotive expectations for cybersecurity and software-update management systems. A small retrofit startup may not immediately fall within the same type-approval responsibilities as a vehicle manufacturer, but these instruments provide an appropriate benchmark for organisational maturity. The key lesson is that cybersecurity is not a one-time penetration test. It is a managed lifecycle involving risk ownership, vulnerability monitoring, incident response, supplier control, and evidence that protections remain effective after updates.

Machine-learning components introduce additional assurance needs. Model files should be versioned and hashed; training and evaluation data provenance should be documented; and performance should be re-evaluated when a model is replaced. Because the current architecture is primarily rules-based and uses facial landmarks rather than an opaque end-to-end classifier, it offers useful interpretability. Nevertheless, the underlying landmark estimator remains a learned component whose limitations must be included in SOTIF analysis and update testing.

G. Olfactory Intervention Safety and Responsible Claims

The olfactory channel differentiates AI Safety Guard from conventional audio-only prototypes, but it also creates the greatest evidentiary and ethical burden. Research on scent and fatigue suggests that selected odours can influence subjective fatigue or driving-related outcomes, yet effects vary by scent, concentration, exposure duration, participant preference, and experimental context (Jiang et al.,

2024). Peppermint aroma has also been associated with changes in driving behaviour in controlled research (Moss et al., 2023). These findings support investigation; they do not establish that a two-second scent pulse will reliably terminate a microsleep or prevent a crash.

A safe research formulation requires materials documentation, concentration control, repeatable atomisation, exposure measurement, and medical screening. The current concept appropriately includes a short maximum spray, cooldown lockout, and physical opt-out. Future hardware should add an independent maximum-on-time circuit and detect a stuck relay where feasible. Consumables should be standardised and tamper-evident. The device should not permit users or fleets to substitute unknown fragrances or irritants, because chemical composition and dose would then become uncontrolled.

Claims should therefore remain narrow. The system may be described as providing an additional salient cue when drowsiness is detected. It should not be described as a treatment for fatigue, a guarantee of wakefulness, or a substitute for rest. Driver-facing instructions should state that any drowsiness alert requires stopping as soon as it is safe and following fatigue-management policy. This wording aligns the innovation with official road-safety advice and prevents an alerting device from encouraging drivers to continue a journey that should be interrupted.

VII. Proposed Validation Programme

A staged programme is recommended so that human trials are not conducted before the system is technically stable. Each phase has explicit entry and exit criteria. Failure in an early phase should result in redesign rather than progression.

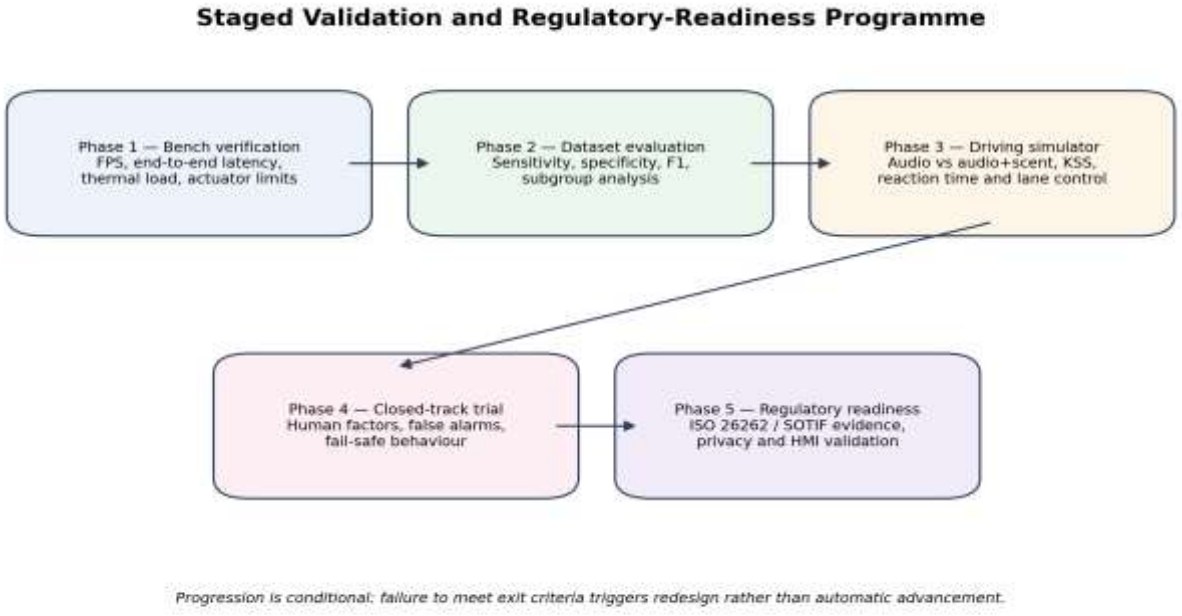

Figure 12. Five-phase validation and regulatory-readiness programme.

A. Phase 1: Bench Verification

Bench testing should characterise performance without human fatigue induction. Test variables include image resolution, lighting, CPU load, thermal state, camera distance, head position, and simultaneous actuation. Instrumented timestamps should measure acquisition-to-feature, feature-to-decision, decision-to-GPIO, GPIO-to-audio, and GPIO-to-scent onset. Electrical measurements should confirm pulse duration, cooldown integrity, power draw, and behaviour after restart.

Exit criteria should include stable operation for an extended duration, no uncontrolled atomiser actuation, documented timing percentiles, recoverable camera failure, and a visible system-health indicator.

Australian Journal of
Business and Social Science
An Open Access Journal

B. Phase 2: Algorithm Evaluation

The detection algorithm should be evaluated in public and locally collected datasets with labelled eye state, yawns, and head pose. Metrics should be reported at frame, event, and driver-session levels. Event-level sensitivity and false alarms per hour are particularly important because high frame-level accuracy can coexist with unacceptable warning frequency.

Subgroup analysis should include lighting, glasses, sunglasses, skin tone, facial hair, headwear, camera angle, and demographic representation where ethically and legally permitted. Ablation studies should compare eye-only, eye+yawn, eye+head, and full fusion. Personalised threshold calibration should be compared with fixed thresholds.

Table 5. Recommended technical evaluation metrics.

Evaluation domain	Primary measures	Reporting requirement
Detection discrimination	Sensitivity, specificity, precision, F1, ROC/PR-AUC	Per channel and fused model; confidence intervals
Operational errors	False alerts per driving hour; missed critical events	Report severity and context, not only aggregate accuracy
Timing	Median, P95, maximum end-to-end latency	Synchronised clocks and defined endpoints
Robustness	FPS, dropped frames, landmark-confidence failures	By lighting, thermal state, camera placement
Fairness	Subgroup sensitivity and false-alarm gap	Pre-specified groups and minimum sample size
Privacy	Bytes written, network transmissions, retention time	Packet capture and storage audit

C. Phase 3: Driving-Simulator Study

A randomised crossover simulator study can compare three conditions: monitoring without alert, audio-only alert, and audio plus optional scent. Participants should complete controlled monotonous drives at a safe time of day under ethics approval. Outcomes should include Karolinska Sleepiness Scale, psychomotor vigilance or reaction-time measures, lane-position variability, steering corrections, braking response, event acknowledgement time, discomfort, and trust.

The primary hypothesis should not be that scent wakes drivers. A more defensible hypothesis is that a bounded dual-sensory alert produces a larger short-term improvement in objective alertness and acknowledgement than audio alone, without worsening lane control or discomfort. The analysis should model repeated measures, order effects, and individual sensitivity. Adverse events and participant withdrawal must be recorded.

D. Phase 4: Closed-Track Field Trial

Only after bench and simulator criteria are met should the system progress to a closed track with trained safety personnel, dual controls where feasible, and predefined abort conditions. The trial should evaluate vibration, road lighting, cabin noise, camera mounting, steering behaviour during alerts, and false alarms under realistic movement. Naturalistic public-road testing would require a further safety case and regulatory approval.

E. Phase 5: Regulatory and Deployment Readiness

The final phase should consolidate requirements traceability, software versioning, cybersecurity, hazard controls, manufacturing consistency, and human factors. A deployment decision should be based on a safety case demonstrating that residual risk is acceptable for a clearly defined use. Post-deployment monitoring should capture device health, false-alarm complaints, adverse scent responses, and software defects without collecting unnecessary driver video.

F. Statistical Analysis and Reporting Plan

The validation programme should be accompanied by a pre-specified analysis plan. For algorithm evaluation, the unit of analysis must be defined at both frame and event levels. Frame-level metrics are useful for debugging but can exaggerate sample size because adjacent frames are highly correlated. Event-level analysis should define a drowsiness episode, matching tolerance, onset time, and minimum separation between episodes. Confidence intervals should be calculated by participant or trip using bootstrapping or hierarchical models rather than treating every frame as independent.

For simulator studies, a repeated-measures crossover design can improve power because each participant experiences each alert condition. The analysis should account for condition order, period effects, baseline sleepiness, time-on-task, and carryover. Mixed-effects models are suitable because they accommodate repeated observations and individual differences. Primary outcomes and time windows should be declared before data inspection. Candidate outcomes include change in lane-position variability, reaction time, steering correction, eye-opening duration, KSS, and time to acknowledgement. Adverse events and premature condition termination must be reported, not excluded as inconvenience.

Sample-size planning should be based on the primary outcome and a minimally important effect, not on a generic rule. A pilot can estimate within-participant variability and feasibility, but confirmatory inference should use a separately powered study. Multiple comparisons should be controlled or clearly distinguished as exploratory. Missing data mechanisms should be documented, particularly when sensor loss is related to eyewear, head pose, or participant characteristics. Reporting only complete and easy-to-track sessions would overstate real-world performance.

For the intervention comparison, non-inferiority or superiority hypotheses should reflect the intended claim. If the aim is to show that audio plus optional scent improves short-term response over audio alone, the primary contrast and effect size must be pre-defined. Acceptance and adverse-response outcomes should be co-primary or key secondary measures because an alert that produces a short response but is frequently disabled cannot deliver practical safety benefit. Results should include participant-level distributions, not only means.

Transparent reporting should include software and hardware versions, camera position, lighting, scent material and dose, threshold configuration, data exclusions, and all planned outcomes. Negative or null results are informative. If scent adds no measurable benefit, or if false alarms remain unacceptable, the evidence should guide redesign rather than be hidden. This research posture is essential for a safety technology whose public value depends on reliable limitations as much as on promising demonstrations.

G. Fairness, Accessibility, and Subgroup Validation

Computer-vision performance can vary across faces and operating conditions. A fairness evaluation should therefore examine whether false-negative or false-positive rates differ by variables that plausibly affect sensing, including skin tone, eye shape, facial hair, headwear, prescription glasses, sunglasses, camera geometry, and lighting. Demographic categories should be collected only where ethically justified and with participant consent. The objective is not to infer identity from facial data; it is to determine whether the system provides comparable safety performance across intended users.
Subgroup evaluation requires adequate sample size and careful interpretation. Small cells can produce unstable estimates, while broad averages can conceal poor performance in a minority condition. Results should report confidence intervals and intersectional limitations. Where a condition such as opaque sunglasses makes visual monitoring impossible, the appropriate response may be an explicit operating limitation and degraded-mode warning rather than a misleading claim of universal coverage. Product documentation should make such boundaries visible to purchasers and drivers.

Accessibility also applies to the alert interface. Drivers with hearing impairment may not perceive an audio alert, while drivers with anosmia or respiratory sensitivity may not benefit from or tolerate scent. A multi-sensory design can support redundancy, but it should not assume that every sensory channel is available. Future versions could examine haptic outputs, provided they are tested

for startle and control interference. Controls and indicators should be usable by people with limited dexterity or colour-vision differences and should not require reading complex information while driving.

Language and cultural factors influence interpretation of warnings and scent acceptability. Fleet deployments should use clear, standardised symbols and locally appropriate voice or tone. User testing should include the intended workforce rather than relying only on technically experienced student developers. Inclusion is a safety requirement: a device that performs well only for the developers or a narrow laboratory sample cannot support universal-deployment claims.

H. Reproducibility, Open Evidence, and Independent Evaluation

Independent evaluation is needed to move beyond project-controlled demonstrations. A reproducibility package should include a bill of materials, wiring diagram, software dependencies, configuration schema, test scripts, and synthetic or de-identified example logs. Safety-sensitive details can be released in a staged manner if unrestricted disclosure would create security risk, but external evaluators must receive enough information to reproduce timing and detection tests. Model and code versions used for each published result should be archived.

A reference test harness can generate timestamped video and simulated GPIO events to measure processing latency independently of human observation. Bench fixtures can verify relay-on duration, cooldown, reset behaviour, startup state, and response to power interruption. Dataset evaluation should provide scripts that recreate confusion matrices and event-level metrics. Where data cannot be shared because of participant privacy, the paper should publish the protocol, variable definitions, codebook, and a controlled-access pathway.

Independent laboratories or university partners can conduct replication under different cameras, lighting, and vehicle cabins. Disagreement between internal and external results should be treated as design information. The 116 ms website example and 0.001-second deck claim demonstrate why reproducibility is valuable: a shared measurement definition can distinguish software decision time, GPIO command time, relay switching time, and physical atomiser onset. Transparent evidence will strengthen both scholarly credibility and future commercial due diligence.

VIII. Discussion

AI Safety Guard occupies a useful design space between basic alarm prototypes and expensive integrated vehicle systems. Its strongest attributes are local processing, interpretable multi-cue logic, low-cost hardware, and an explicit attempt to move from passive detection to active intervention. The architecture is compatible with current edge-computing research and responds to privacy concerns that arise when driver video is uploaded to external servers.

The tri-factor model is technically sensible because eye closure, yawning, and head pose represent different manifestations of drowsiness. Fusion can reduce dependence on any single cue, especially when implemented with persistence and confidence checks. However, fusion does not automatically improve safety; poorly tuned additional channels can increase false alarms. The contribution lies in the testable architecture, not in an assumption that more sensors always produce better classification.

The scent channel is the most distinctive and the most uncertain element. It may provide a salient second cue and reduce habituation to audio, but it introduces exposure, acceptability, and human-factors questions. The correct research strategy is controlled comparison with audio-only intervention, not deployment based on anecdotal alertness. The physical opt-out and cooldown described in the project are appropriate early controls and should be strengthened with independent hardware timing.

The prototype also illustrates why performance claims must be defined precisely. A claim of 0 ms network latency is directionally true for offline processing but is not a measure of end-to-end warning time. A claim of 0.001 s actuator latency may represent electrical switching rather than perception. A false-positive rate below 1% is uninterpretable without the number of frames, events,

and driving hours. Publication-ready engineering requires denominators, timestamp boundaries, uncertainty, and reproducibility.

From an educational innovation perspective, the project demonstrates the value of integrating computer vision, edge computing, electronics, user controls, and safety reasoning. Its award recognition indicates that judges perceived both technical integration and societal relevance. The next step is to convert this demonstrator into a research platform whose logs, test fixtures, protocols, and version control can support independent replication.

The paper's safety-by-design framing is deliberately conservative. Driver monitoring is a safety-related function, and a prototype that generates physical outputs in a vehicle must be evaluated for both intended and unintended behaviour. The device should never promise that the driver can continue safely after an alert. Its safest role is to detect concerning cues, issue a graded warning, and escalate to a safe-stop recommendation.

A. Theoretical Contribution

The study contributes a layered model of drowsiness-intervention systems: perception, temporal inference, safety control, human response, and governance. Much prototype literature focuses on perception accuracy and treats the alarm as trivial. This paper shows that intervention and governance are independent research objects. Detection accuracy does not determine whether the alert is acceptable, and privacy does not follow automatically from edge deployment.

The design also illustrates an evidence hierarchy for student and startup prototypes. Design coherence and bench demonstration are legitimate contributions when clearly bounded. They become problematic only when converted into unsupported claims of medical, behavioural, or crash-prevention efficacy. The proposed hierarchy can be applied to other edge-AI safety systems.

B. Practical Implications

For developers, the immediate priorities are instrumented latency measurement, hardware-enforced atomiser limits, confidence-aware degraded modes, and reproducible test scripts. For researchers, the priority is an ethics-approved simulator study with objective driving outcomes. For institutions and competition programmes, the project demonstrates why mentoring should include evidence discipline, product safety, and regulatory literacy alongside software development.

For fleet operators, the system should not be used as a disciplinary surveillance device or a substitute for fatigue-management schedules. Any trial should involve drivers, safety representatives, privacy officers, and occupational health expertise. Data governance should prohibit secondary use of facial or event data without clear authority.

C. Deployment Scenarios and Operational Boundaries

The appropriate system configuration depends on deployment context. For a private consumer vehicle, the priority may be simple installation, local privacy, and a non-persistent session profile. For commercial freight or rideshare fleets, the device must integrate with fatigue-management policy, maintenance schedules, worker consultation, and incident response. A fleet dashboard, if developed, should prioritise device health and aggregate risk trends rather than continuous individual surveillance. For training or research simulators, greater data collection may be acceptable under consent, but results should not be assumed to transfer directly to public-road operation.

Operational boundaries should be documented as part of intended use. The prototype is not a medical device, not an autonomous-driving function, and not a substitute for a fit-for-duty assessment. It does not control steering or braking. It should not be relied upon when the camera cannot see the face, when the driver wears eyewear that blocks eye tracking, or when installation falls outside the validated geometry. It should not encourage extended driving after an alert. These limitations should be incorporated into the interface, manual, training, and marketing materials.

Maintenance is part of safety performance. Camera lenses require cleaning; mounts require inspection; scent consumables require controlled replacement; speakers and buttons require self-test;

and software versions require management. A system that operates correctly at installation but gradually loses camera alignment or actuator function can create false reassurance. Fleet trials should therefore evaluate maintainability and diagnostic coverage, not only detection accuracy.

The low-cost goal remains valuable, but cost comparisons should include enclosure, automotive power protection, installation, certification, maintenance, consumables, cybersecurity support, and end-of-life disposal. A sub-$100 laboratory bill of materials demonstrates accessibility of prototyping, not the final cost of a compliant product. Clear separation of prototype cost from deployment cost will improve the quality of future commercial and social-impact claims.

D. Translational Research and Commercialisation Roadmap

Translation from prototype to deployable products should be milestone-based. The first milestone is a reproducible engineering prototype with version-controlled hardware, calibrated timing, fault supervision, and a documented hazard log. The second is an algorithm-validation build tested on held-out datasets and a representative cabin test matrix. The third is a human-factors research device with ethics-approved exposure controls and medical screening. The fourth is a closed-track product supported by electrical, electromagnetic, environmental, mechanical, cybersecurity, and software-update evidence. Only after these stages should public-road or commercial deployment be considered.

Commercialisation decisions should preserve the project's safety rationale. A business model based on selling event data or continuous worker analytics could conflict with the privacy-first value proposition and reduce driver trust. Alternative models include device sales, fleet maintenance subscriptions, or safety-program support in which only minimal device-health and alert summaries are retained. Procurement materials should state evidence level, intended use, limitations, and validation status, avoiding comparison claims that have not been tested.

Partnerships will be necessary. Automotive engineers can address electrical and mechanical integration; sleep and human-factors researchers can design valid fatigue protocols; occupational-health specialists can advise fleet use; toxicologists or exposure scientists can evaluate scent; and regulators or conformity-assessment bodies can clarify approval pathways. The university context is well suited to this multidisciplinary transition because it can separate exploratory research from commercial pressure while building an auditable evidence base.

The INCITE award provides meaningful external recognition of innovation and communication, but it is not a substitute for safety validation. Used responsibly, the recognition can attract collaborators and resources for the staged programme described in this paper. The strongest commercial narrative is therefore not that the project has already solved drowsy driving, but that it offers a distinctive, low-cost, privacy-preserving research platform with a credible plan for testing active multi-sensory intervention.

E. Broader Research Contribution

Beyond the specific prototype, the study offers a general model for responsible development of AI-enabled physical interventions. Many student and startup projects combine machine perception with an actuator, yet evaluation often focuses on model accuracy while under-specifying the consequences of false activation, delayed activation, or silent failure. The layered framework in this paper links perception, temporal inference, safety control, physical output, human response, and governance. This structure can be applied to workplace safety, assistive technology, health monitoring, and other edge-AI systems that act on people or environments.

The paper also demonstrates the importance of treating privacy and safety as co-requirements. Local inference reduces exposure of facial imagery and network dependency, but it does not automatically guarantee safe behaviour. Conversely, extensive cloud analytics may support remote monitoring but create privacy and cybersecurity risk. The appropriate architecture must balance sensing performance, latency, transparency, data minimisation, maintenance, and accountability. AI Safety Guard provides a concrete case through which these trade-offs can be examined.

Finally, the evidence hierarchy offers a publication framework for prototype research. Design papers can make a legitimate contribution when they clearly distinguish artefact description, bench observation, proposed validation, and demonstrated efficacy. This avoids both extremes: dismissing early-stage engineering because it lacks a clinical trial, and overstating a demonstration as proof of social impact. The approach enables rigorous scholarship while the technology remains at a formative stage.

IX. Limitations

The study is limited by its reliance on project-controlled documentation and publicly visible prototype material. Source code, raw logs, calibration data, electrical schematics, and complete test records were not independently audited. The website and deck contain differing latency figures, and several claimed metrics lack denominators. Accordingly, the paper does not independently verify accuracy, false-positive rate, or response time.

No human-participant data are reported. The analysis cannot determine whether audio plus scent improves alertness, whether scent causes adverse effects, or whether drivers accept the intervention. The literature on olfactory cues is supportive but heterogeneous and does not validate this implementation.

The regulatory analysis is a research-oriented benchmark rather than legal advice or a conformity assessment. Requirements vary by jurisdiction, vehicle category, intended use, and deployment model. Standards referenced in the paper require access to full normative documents and qualified safety engineering.

Finally, visual drowsiness cues are proxies. A driver may be cognitively impaired without obvious eye or head changes, and visible behaviours can occur without dangerous sleepiness. Future systems may benefit from vehicle signals or non-contact physiological sensing, but every additional modality introduces cost, privacy, and integration trade-offs.

X. Conclusion

AI Safety Guard is a coherent edge-AI prototype that integrates camera-based drowsiness cues, temporal fusion, local processing, and bounded multi-sensory warning. Its architecture addresses a genuine road-safety problem and demonstrates a privacy-conscious alternative to cloud-dependent driver monitoring. The prototype's innovation is strengthened by interpretable tri-channel logic, physical scent opt-out, cooldown lockout, manual acknowledgement, and minimal event logging.

The current evidence supports feasibility, not road-readiness. Reported frame rate, false-positive rate, and latency require reproducible measurement, and the olfactory intervention requires controlled human-factors evaluation. The paper resolves this boundary by presenting a five-phase validation programme and a preliminary safety case aligned with functional safety, SOTIF, privacy, cybersecurity, and regulatory principles.

Progression should be conditional: bench stability before participant testing, participant evidence before closed-track trials, and a documented safety and cybersecurity case before public-road deployment. The expanded analysis in this paper adds requirements for calibration, confidence-aware degraded modes, installation engineering, privacy governance, software-update assurance, subgroup fairness, reproducibility, and statistically defensible evaluation. Under this evidence-bounded pathway, AI Safety Guard can develop from an award-winning educational prototype into a credible multidisciplinary research platform for privacy-preserving driver drowsiness warning and responsible edge-AI actuation.

Declarations

Funding: No external research funding is reported for this manuscript.
Conflict of interest: The author is associated with the supervision and development context of the AI Safety Guard project. This relationship is disclosed because the manuscript analyses project-controlled documentation.

Ethics: The present manuscript reports no human-participant experiment. Any future simulator or scent-exposure study will require prior institutional ethics approval, informed consent, adverse-event procedures, and appropriate medical screening.

Data availability: The public project website, the thirteen-page project deck, and the twenty-three-slide AI Drowsiness Detection System presentation are the principal artefact sources. No participant-level dataset is analysed in this paper.

Author responsibility: The author takes responsibility for the interpretation, evidence boundaries, and manuscript content.

Acknowledgements

The author acknowledges Pattricia Ruth Das for product development and Dr Reza Ghanbarzadeh for co-supervision of the project. Award recognition by the 35th INCITE Awards is acknowledged as external recognition of innovation, not as scientific validation.

References

AI Safety Guard. (2026). SafetyGuard v1.2: A student-built, lab-tested edge-AI drowsiness-detection prototype. https://aisafetyguard.netlify.app/

AI Safety Guard Project Team. (2026a). AI Safety Guard: Multi-sensory edge AI for driver drowsiness detection [Unpublished project deck].

AI Safety Guard Project Team. (2026b). AI Drowsiness Detection System [Unpublished technical presentation].

Albadawi, Y., AlRedhaei, A., & Takruri, M. (2022). A review of recent developments in driver drowsiness detection systems. Sensors, 22(5), 2069. https://doi.org/10.3390/s22052069

Albadawi, Y., AlRedhaei, A., & Takruri, M. (2023). Real-time machine learning-based driver drowsiness detection using visual features. Journal of Imaging, 9(5), 91. https://doi.org/10.3390/jimaging9050091

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. International Journal of Neuroscience, 52(1-2), 29-37. https://doi.org/10.3109/00207459008994241

Australian Government. (2026, February 21). Stay alert, stay alive: National Driver Fatigue Week targets preventable road tragedies. https://minister.infrastructure.gov.au/chisholm/media-release/stay-alert-stay-alive-national-driver-fatigue-week-targets-preventable-road-tragedies

Chew, Y. X., Abdul Razak, S. F., Yogarayan, S., & Sayed Ismail, S. N. M. (2024). Dual-modal drowsiness detection to enhance driver safety. Computers, Materials & Continua, 81(3), 4397-4417. https://doi.org/10.32604/cmc.2024.056367

European Commission. (2021). Commission Delegated Regulation (EU) 2021/1341 of 23 April 2021 supplementing Regulation (EU) 2019/2144 with detailed rules concerning driver drowsiness and attention warning systems. https://eur-lex.europa.eu/eli/reg_del/2021/1341/oj

European Parliament and Council. (2019). Regulation (EU) 2019/2144 on type-approval requirements for motor vehicles and systems as regards their general safety and the protection of vehicle occupants and vulnerable road users. https://eur-lex.europa.eu/eli/reg/2019/2144/oj

Hassan, O. F., Ibrahim, A. F., Gomaa, A., Makhlouf, M. A., & Hafiz, B. (2025). Real-time driver drowsiness detection using transformer architectures: A novel deep learning approach. Scientific Reports, 15, 17493. https://doi.org/10.1038/s41598-025-02111-x

INCITE Awards. (2026). Winners of the 35th INCITE Awards announced. https://www.inciteawards.org.au/latest-news/winners-of-the-35th-incite-awards-announced/

International Organization for Standardization. (2018). ISO 26262-2:2018 Road vehicles - Functional safety - Part 2: Management of functional safety. https://www.iso.org/standard/68384.html

International Organization for Standardization. (2022). ISO 21448:2022 Road vehicles - Safety of the intended functionality. https://www.iso.org/standard/77490.html

International Organization for Standardization. (2023). ISO 24089:2023 Road vehicles - Software update engineering. https://www.iso.org/standard/77796.html

International Organization for Standardization & SAE International. (2021). ISO/SAE 21434:2021 Road vehicles - Cybersecurity engineering. https://www.iso.org/standard/70918.html

Jiang, X., Muthusamy, K., Chen, J., & Fang, X. (2024). Scented solutions: Examining the efficacy of scent interventions in mitigating driving fatigue. Sensors, 24(8), 2384. https://doi.org/10.3390/s24082384

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska Sleepiness Scale against performance and EEG variables. Clinical Neurophysiology, 117(7), 1574-1581. https://doi.org/10.1016/j.clinph.2006.03.011

Maior, C. B. S., Moura, M. J. C., Santana, J. M. M., & Lins, I. D. (2020). Real-time classification for autonomous drowsiness detection using eye aspect ratio. Expert Systems with Applications, 158, 113505. https://doi.org/10.1016/j.eswa.2020.113505

Makhmudov, F., Turimov, D., Xamidov, M., Nazarov, F., & Cho, Y.-I. (2024). Real-time fatigue detection algorithms using machine learning for yawning and eye state. Sensors, 24(23), 7810. https://doi.org/10.3390/s24237810

Moss, M., Ho, J., Swinburne, S., & Turner, A. (2023). Aroma of the essential oil of peppermint reduces aggressive driving behaviour in healthy adults. Human Psychopharmacology: Clinical and Experimental, 38(2), e2865. https://doi.org/10.1002/hup.2865

National Highway Traffic Safety Administration. (2026). Drowsy driving: Avoid falling asleep behind the wheel. https://www.nhtsa.gov/risky-driving/drowsy-driving

Owen, V., & Surantha, N. (2025). Computer vision-based drowsiness detection using handcrafted feature extraction for edge computing devices. Applied Sciences, 15(2), 638. https://doi.org/10.3390/app15020638

United Nations Economic Commission for Europe. (2021a). UN Regulation No. 155 - Cyber security and cyber security management system. https://unece.org/transport/documents/2021/03/standards/un-regulation-no-155-cyber-security-and-cyber-security

United Nations Economic Commission for Europe. (2021b). UN Regulation No. 156 - Software update and software update management system. https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update

United Nations Economic Commission for Europe. (2026a, March 12). New UN regulation to prevent fatigue-related crashes. https://unece.org/media/transport/Road-Safety/press/412103

Yang, Y., Wu, X., Wang, L., Easa, S. M., & Zheng, X. (2024). Evaluating acceptance of novel vehicle-mounted perfume automatic dispersal device for fatigued drivers. Applied Sciences, 14(11), 4580. https://doi.org/10.3390/app14114580

Appendix A. Proposed Reproducible Bench-Test Matrix

Table A1. Bench-test matrix for reproducible prototype verification.

Test area	Conditions	Measures	Minimum reporting
Camera/inference	Daylight, low light, glare, night, shadow transition	FPS, dropped frames, landmark confidence, detection events	30 min per condition; report distribution
Thermal endurance	Continuous operation in enclosed cabin temperature range	CPU temperature, throttling, FPS, crashes	At least 4 h per configuration
Eye closure	Natural blink, deliberate closure, partial closure, eyewear	Event sensitivity, false alarms, time-to-detect	Scripted repetitions by participant/mannequin where ethical
Yawn/speech	Yawns, speech, singing, coughing, mouth movement	Yawn precision and event-level F1	Time-stamped video labels
Head pose	Head turn, mirror check, downward glance, head drop	Pose error and critical-event errors	Ground-truth angle where feasible

Test area	Conditions	Measures	Minimum reporting
Actuation	Audio only, scent enabled, relay fault, restart	Onset latency, pulse duration, lockout, safe state	Oscilloscope/current sensor plus video/audio timestamp
Privacy	Normal use, debug mode, update mode	Network packets, files written, retention	Packet capture and storage diff

Appendix B. Minimum Event Log Schema

A privacy-minimised event record should contain: device software version; anonymised session identifier; monotonic and wall-clock timestamp; camera health; frame rate; face/landmark confidence; calibrated thresholds; eye-channel state; yawn-channel state; head-pose state; fused risk reason; alert state; scent-enabled flag; acknowledgement time; cooldown state; and fault codes. It should not contain raw video, face images, audio recordings, or personal identity unless a separately approved research protocol requires them.

Appendix C. Recommended Simulator Outcomes

Table C1. Proposed outcome measures for a controlled

Domain	Example measures	Purpose
Primary safety	Standard deviation of lateral position; lane departure; reaction time	Objective effect on driving control
Drowsiness	Karolinska Sleepiness Scale; PVT/reaction task; eye-closure events	Short-term alertness response
Intervention	Acknowledgement time; recovery duration; repeat warning	Human response to alert
Acceptability	Comfort, startle, irritation, scent acceptability, trust	Human factors and adoption
Safety events	Steering disturbance; panic response; adverse scent symptoms	Adverse-event assessment